DOCUMENT RESUME

ED 222 074 FL 013 251

AUTHOR Cziko, Gary A.
TITLE Approaches to the Evaluation of Bilingual Education: An International Perspective. Professional Papers CZ-1.
INSTITUTION National Center for Bilingual Research, Los Alamitos, Calif.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jun 82
NOTE 24p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bilingual Education; *Bilingual Education Programs; Educational Assessment; Elementary Education; *Evaluation Methods; Foreign Countries; *Immersion Programs; *Program Evaluation
IDENTIFIERS Canada; Nigeria; Sudan

ABSTRACT

The evaluation of bilingual education programs is complicated by such factors as the diversity of evaluation methodologies and program goals and the reliability of instruments for minority language students. Three bilingual program evaluations in foreign countries are described in terms of their different contexts and approaches in order to raise issues about bilingual education program evaluation. The programs evaluated were the St. Lambert French immersion program in Canada, the Yoruba 6-year primary project in Nigeria, and the local language literacy training project in the southern Sudan. Based on these evaluation experiences, the strengths and weaknesses of quantitative and qualitative evaluation methods are discussed. A combination of quantitative and qualitative evaluation methods is suggested as a means of maximizing the strengths of each approach. However, it is important that such a combined approach be carefully designed. (RW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222074

APPROACHES TO THE EVALUATION OF BILINGUAL EDUCATION: AN INTERNATIONAL PERSPECTIVE

Gary A. Cziko

June, 1982

[redacted] CZ-1

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCBR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL013251

national center for bilingual research, 4665 lampson avenue, los alamitos, CA 90720, telephone (213) 598-0481

The material in this publication was prepared under Cooperative Agreement 00-CA-80001 from the National Institute of Education, Department of Education. Its contents do not necessarily reflect the views of the National Institute of Education or of any other agency of the United States of Education or of any other agency of the United States government, and no official endorsement should be inferred.

Approaches to the Evaluation of Bilingual Education: An International Perspective

Gary A. Cziko

University of Illinois

APPROACHES TO THE EVALUATION OF BILINGUAL EDUCATION: AN INTERNATIONAL PERSPECTIVE

Gary A. Cziko

Introduction

The problem of evaluating bilingual education programs in the U.S. is an exceedingly complex one. First, there is the problem of choosing an appropriate methodology for conducting evaluations which must take into account the debate between quantitative and qualitative approaches to evaluation. Second, there is a diversity of goals for bilingual education in the U.S. with some programs attempting to transition students into all-English programs as quickly as possible while others attempt to maintain or restore knowledge of the students' first language and culture. Third, one must take into account the possibility that measures commonly used to assess academic achievement, language proficiency, and attitudes may not possess adequate reliability and/or validity and may be seriously biased against children who have not had much contact with the majority language and culture of this country. Fourth, student success within bilingual education programs appears to be influenced not only by the type and quality of their educational program, but also by psychological and sociolinguistic factors operating within the context of the classroom, school, and community. Finally, evaluators working within the context of bilingual programs must have adequate knowledge of the language and culture of the groups with which they are working.

It is no wonder, then, that there is so much debate, controversy, and disagreement concerning both the appropriate methodology for evaluating bilingual programs and the interpretation and implications of the many bilingual education program evaluations that have been conducted. To adequately address all of these issues would require a major effort and may well be beyond the ability of any one person; it is certainly beyond the scope of this paper. The objective of this paper

is considerably more modest. It will attempt to bring an international perspective to bear on problems of concern to bilingual education evaluators and educators in this country.

In contrast to most fields of education in which the U.S. is at the forefront in research, bilingual education has been one area where American researchers have spent considerable time examining the success of other countries with what is both a very old and a very new approach to education. Although I am now actively involved in bilingual education evaluation and research in the U.S., I have been fortunate to have been involved in the evaluation of a number of bilingual education programs outside this country. In this paper, I will briefly describe three bilingual education projects I have evaluated--one in Canada and two in Africa--which are particularly interesting because they took place in very different contexts and were evaluated using quite different approaches, approaches reflecting both changes in the field of educational evaluation and my own development as an evaluator. I will use these three projects and their evaluations to raise a number of basic issues relating to the evaluation of bilingual education programs. I will then make some tentative conclusions and suggestions concerning the evaluation of bilingual education in the U.S.

Canada: French Immersion Programs

One of the most influential evaluations of a bilingual education program has been that of the French immersion begun in 1965 in St. Lambert, a suburb of Montreal (Lambert and Tucker, 1972). The purpose of this program is to allow English-Canadian children to acquire functional bilingualism in French and English. This is done by teaching all subjects in French until Grade 2 or 3, at which time English language arts are introduced for the first time. While virtually all evaluations of the original St. Lambert program, as well as other French immersion programs throughout Canada, have been favorable, it is

interesting to take a close look at the approach used to evaluate the original St. Lambert project.

The approach used by Lambert and Tucker (1972) can perhaps be described as an experimental, quantitative approach to educational evaluation. This is characterized by the selection of experimental and control groups, attempts to equate the experimental and control groups before the beginning of the program, the systematic administration of a large number of measures of academic achievement, language proficiency, language use, and attitudes, followed by statistical tests of significance between the experimental and control groups. In short, all data were quantified and all judgments of the impact of the program were based on statistical comparisons between the experimental and control groups. The only qualitative data to appear in the original evaluation appear as an appendix in Lambert and Tucker's book which briefly describes the class activities of the experimental group from kindergarten through Grade 5.

In spite of the fact that the evaluation approach used by Lambert and Tucker would not be appropriate for the great majority of bilingual education programs in the U.S., this and subsequent evaluations of French immersion programs in Canada have been taken by many in this country as a standard for evaluating bilingual education programs in the U.S. There appear to be a number of reasons for this. First, the general findings of the evaluation of the original St. Lambert project have been replicated many times throughout Canada (see Genesee, 1976) and it has become a generally accepted fact that these programs reliably result in favorable outcomes among English-Canadian children. It has also become quite clear that it is the French immersion program itself, and not other confounding factors, which is responsible for the increased French-language proficiency of children participating in these programs. Also, the successful outcome of the French immersion programs has appeared to have raised relatively little opposition in Canada due to its voluntary nature, low cost, and acceptability to both French- and

English-Canadians alike. Due to these factors, French immersion programs have enjoyed increasing popularity and can now be found in every major city and province throughout Canada. It should be noted, however, that this most prevalent form of bilingual education in Canada has been designed for English-speaking children, which form the majority linguistic group in Canada, and that minority groups which speak a language other than French or English enjoy no legal rights to the use of their native language in public schools throughout Canada. The French immersion programs in Canada have essentially shown that middle-class language majority children can take a heavy dose of a second language and learn it remarkably well without detrimental effects to their first-language development. This finding certainly has important educational implications south of the Canadian border, but not necessarily for linguistic minority children in the U.S.

While the results of the French immersion evaluations have served to extend bilingual education in Canada, these same results have been used repeatedly by critics of U.S. bilingual education efforts in the U.S. as a rationale for eliminating bilingual education in this country. In spite of the fact that the Canadian researchers have repeatedly warned that their findings are not generalizable to linguistic minority children in the U.S., these evaluations have been repeatedly used as evidence that children can be immersed in a second language at school with no ill effects to their linguistic and academic development (see Baker and deKanter, 1981; Epstein, 1977, pp. 53-54). This has been done in spite of the fact that there are probably more differences than similarities between English-Canadian children in French immersion programs and linguistic minority children in the U.S. who are submerged into all-English educational programs.

In addition to the danger of overgeneralizing the Canadian evaluation results, another danger to bilingual education in the U.S. lies in the experimental, quantitative approach to evaluation used in the Canadian studies. This approach to evaluation seems to be commonly

regarded as the only "scientific" way of demonstrating the effectiveness of bilingual education programs, in spite of the fact that such an approach is usually impossible to apply to bilingual education programs in the U.S., since it is illegal to keep eligible children out of bilingual programs to create control groups. Another problem with this approach to evaluation is its tendency to rely heavily on quantitative outcome measures while putting relatively little emphasis on describing the context of the community, school, and classroom in which the program takes place. Although it is clearly important to measure the outcomes of bilingual education programs, the great variation in the way these programs are implemented makes the collection of outcome data of little practical use without a detailed description of the program and its context. We already know that there are both effective and ineffective bilingual education programs in the U.S. (see Troike, 1978). What we need to know are the factors which differentiate effective from ineffective programs. Evaluations which give little attention to what goes on in the classroom and to what the children, parents, teachers, and administrators think and feel about the program would not appear to offer much useful information concerning why some programs are effective and others are less so.

Nigeria: The Yoruba Six-Year Primary Project

Prior to 1970, both Yoruba (for Primary 1 through 3) and then English (for Primary 4 through 6) were used as media of instruction in the six-year primary education of all Yoruba-speaking children in the western part of Nigeria. The Yoruba Six-Year Primary Project (Afolayan, 1976) was initiated in 1970 in an attempt to devise a program which would make the primary education of these children more effective and meaningful by using Yoruba as the sole medium of instruction for the first six years of school. To test the effects of the exclusive use of Yoruba as the medium of instruction, a research project was initiated. Experimental and control classes were set up at St. Stephen's "A" School in Ile-Ife. Early in the implementation of the project, however, the

project administrators found what they believed to be serious defects in the primary school curriculum and so took on the job of creating a new curriculum incorporating the subjects of Yoruba, English, science, social and cultural studies, and mathematics. Both the St. Stephen's experimental and control groups have made use of this new curriculum. In addition, a specialist teacher of English was used to provide English instruction for the experimental class while the usual classroom teacher provided English instruction for the control class.

In 1973, the project was expanded to include ten additional "proliferation" schools, eight of which were to use Yoruba as the sole medium of instruction (the proliferation group) while the remaining two were to follow the usual pattern of three years of Yoruba followed by three years of English (the proliferation control group). A comprehensive evaluation of the Yoruba Primary Project was initiated in 1976 with the testing of academic achievement, Yoruba- and English-language skills, and intelligence of Primary 3 children in the St. Stephen's and proliferation experimental and control classes, as well as children in selected traditional schools. This large-scale, longitudinal evaluation was to continue until these children had completed primary school to determine the effect of using Yoruba as the exclusive medium of instruction, the impact of the new curriculum materials, and the effectiveness of using a specialist teacher for the teaching of English.

The evaluation design that evolved for this project was primarily quantitative, relying on a large number of tests that were administered to project and control schools in both urban and rural settings and analysis of covariance to compare test performance controlling for differences in socioeconomic status. In retrospect, it is in some ways surprising that such an approach was used, since it involved the construction of a large number of tests (few tests in the Yoruba language were available) and sophisticated data analysis techniques (e.g., analysis of covariance) made possible only by the presence of

modern computing facilities at the University of Ife. (It should be mentioned that initially the project staff felt that a systematic, quantitative evaluation organized by an outsider was not necessary, since they were already personally convinced of the success of the project. It was only after continued pressure from the Ford Foundation, which had provided substantial financial support for the project, that the staff agreed to evaluate the project.) The evaluation of this project produced fairly favorable results, showing that for fostering academic achievement, Yoruba is as effective or more effective than English as a medium of instruction throughout all six years of primary school in the Yorba-speaking areas of western Nigeria (see Ojerinde, 1979). However, as in the original evaluation of the St. Lambert French immersion program in Montreal, apparently no information has been collected on the content and form of classroom activities in either project or control schools, and continued evaluation of the progress of participating students needs to be done as they continue through secondary school, where all instruction takes place in English.

The Southern Sudan: The Local Languages Literacy Training Project

The Southern Regional Ministry of Education of the Sudan, in cooperation with the Summer Institute of Linguistics, is currently involved in a comprehensive project designed to teach literacy skills to elementary school pupils in the Southern Sudan, using 9 of the 53 or so local languages of the region. The project involves the development, production, and dissemination of materials in the local languages, as well as the training of teachers in the use of these materials.

The problems encountered in developing and implementing an evaluation of this project were considerable, due to the fact that the Southern Sudan is one of the poorest and least developed areas of the world. Some of these problems were the unavailability of testing instruments, the difficulty of communication and travel in the area, and difficulty in locating educated native speakers of the local languages

to aid in test construction and data analysis. Nevertheless, the first impact evaluation of the project took place in November/December 1980, coinciding with the in-class use of trial editions of Primary 1, 2, and 3 materials in Bari, Lotuho, Dinka, and Ndogo. Of these four languages, Bari and Lotuho were selected, due to the relative accessibility of the Bari and Lotuho schools from Juba, the regional capital. During a visit to the region in June 1980, the author visited several rural Bari and Lotuho primary schools, some using the new materials, some not. Four of these schools were selected for inclusion in the impact evaluation--a Bari school using the project's Bari materials for Primary 1 and 2, a comparison Bari school not using the project's materials, a Lotuho school using the project's Lotuho materials for Primary 1 and 2, and a Lotuho school not using the project's Lotuho materials. All four schools provided at least basic necessary facilities, i.e., shelter, blackboards, chalk, paper and writing instruments.

Three general types of information were collected from the four schools. First, general background information was collected on the size of each school (enrollment at each grade), the curriculum (subjects taught at each grade, by whom, using what materials), the teachers (education, teaching experience, subjects taught), and the Primary 2 pupils (name, mother tongue, age). Second, information was collected on the actual teaching activities of the Primary 2 vernacular teacher at each school. These data were obtained by tape recording and taking notes on a complete vernacular reading lesson which the author attended, assisted by an educated adult speaker of the vernacular who was fluent in English and who was able to provide the author with explanations and interpretations of the class activities. Finally, information was collected on the actual reading performance of Primary 2 pupils in each of the four Primary 2 classes. This was obtained by adminstering a group test of word recognition to each class and by tape recording performance on individual tests of oral reading and reading comprehension. The oral reading test consisted of four parts: (a) A list of ten words included in the project materials; (b) a list of ten

words not contained in the project materials; (c) a short story of approximately 50 words containing all the words in the two lists; and (d) five comprehension questions based on the story. Each pupil was asked to read aloud the lists and story and to answer orally the questions pertaining to the story. In addition, each pupil was asked to give his or her reasons for attending school and for wanting to learn to read.

The outcome data collected via quantitative test instruments clearly showed that pupils in both the project and comparison classes were having difficulty learning to read. For example, on the story reading test described above, the project pupils tested could only read a mean of 47% of all the words of the story. While this in itself is informative, it is the process data collected in the reading classrooms which give us clues as to why pupils were experiencing difficulties. Virtually all of the reading activities involved mechanical repetition or recitation of letters, words, phrases, sentences, or stories either presented in the pupils' materials or printed on the board, with only a very few instances of activities which required pupils to attempt to comprehend what they read.

Also, the analysis of the errors made by some of the Bari project pupils in answering the class comprehension questions was revealing. Since many of the comprehension questions could be answered by simply repeating an appropriate sentence or part of a sentence from the story, it was often not clear whether the pupils were actually understanding the stories or simply memorizing them from repeatedly hearing them read aloud by the teacher and by the class. However, three pupils during the Bari project literacy lesson began their answers to comprehension question with the word a, which means roughly and then and is often used at the beginning of a non-initial sentence of a Bari story for the purpose of text cohesion. Answering an oral question with a sentence beginning with a is not appropriate (in fact the teacher vigorously corrected these pupils) and seems to indicate that these pupils had in

fact memorized the story and went so far as to violate some basic discourse rules of spoken Bari to use what they had memorized to answer the question.

Fortunately, the information obtained from an analysis of the reading materials, the classroom observations, and the reading test results have led to a number of recommendations concerning materials and teaching techniques which, if followed, we feel will have a positive impact on the acquisition of literacy skills in the vernacular languages as well as in English and Arabic (see Cowan, 1980; Cziko, in press). Future planned evaluations of the project will allow us to determine the feasibility and impact of such changes on the ac uisition of literacy skills in the Southern Sudan.

Implications for the Evaluation of Bilingual Education Programs

In this final section, I will attempt to draw implications from the experiences described above, focusing on the distinction between quantitative and qualitative approachs to evaluation and their particular strengths and weaknesses when applied to evaluating bilingual education programs in the U.S. and abroad. This section will deal with (a) strengths and weaknesses of quantitative approaches to evaluation of bilingual education, (b) strengths and weaknesses of qualitative approaches to evaluation of bilingual education, and (c) a rationale for combining aspects of both quantitative and qualitative evaluation approaches for evaluating bilingual education programs.

Quantitative Evaluation Methods. Quantitative evaluation methods have a number of important strengths, not the least of which is the respect that quantitative, experimental research methods still enjoy among many educational researchers and educational policy planners. The primary advantage of this evaluation approach is that, when properly followed, it can provide convincing evidence that a particular bilingual education program is or is not having a measurable impact on

participating students. In addition, it permits isolation of the effects of the bilingual education program from the possible effects of other competing, potentially confounding factors. This approach also has the advantage of producing results which can be summarized fairly easily, a distinct advantage when a number of different program evaluations are being considered to provide evidence for planning educational policy. However, the design requirements of a properly implemented quantitative evaluation, especially one following an experimental or quasi-experimental design, are such that this approach is seldom feasible for evaluating bilingual education programs in the U.S.

These designs require the random assignment of students to the experimental and control groups or, at the least, assurance that the two groups do not differ systematically on any factors other than the differences in educational program which might influence the evaluation results. Anyone familiar with bilingual education in the U.S. knows that random assignment of children who are eligible for bilingual education to bilingual and non-bilingual programs is usually not possible and, in fact, would likely be in violation of both Federal and state regulations. Also, eligible children who do not participate in bilingual education programs do not do so for two principal reasons--either their parents have decided, for whatever reason, that they do not want their children to receive bilingual instruction, or the children are part of a language group represented by less than 20 children in the school district.[1] Therefore, these children would appear to differ in important and relevant ways from children of the same language group who

---

[1]For a summary of state legislation on the number of students necessary to "trigger" a bilingual education program in a district, see Gray, 1981.

are receiving bilingual education and would not appear to comprise a proper control group.[2]

In addition to the problem of selecting students for the bilingual and non-bilingual programs for the purpose of conducting a quantitative evaluation, program variables other than the presence or absence of bilingual education may also confound the evaluation results (see Baker and deKanter's, 1981, criticism of the McConnell study). In fact, in one analysis of the results of the Yoruba evaluation described above, it was found that there were statistically significant differences on test performance among four classes which were all part of the same treatment group (Cziko and Ojerinde, 1976). For these and other reasons, it is extremely difficult, if not impossible, to undertake a quantitative evaluation of a bilingual education program using an evaluation design above reproach by either proponents or critics of bilingual education. It should not be surprising, therefore, that of the several hundred evaluations reviewed by Baker and deKanter (1981) only 28 were found to be free of serious methodological problems typical of quantitative evaluations of bilingual education programs.

---

[2] It is interesting to note that failure to comply with the requirements of experimental or quasi-experimental evaluation designs has been the principal crticism of quantitative evaluations of bilingual education in the U.S. by both critics and proponents of bilingual education. It seems that when evalutions find bilingual education to be ineffective, proponents of bilingual education are quick to point out the non-equivalence of the bilingually educated and all-English educated groups (see, e.g., Gray's, 1978, criticism of the AIR evaluation by Danoff (1978)) and, conversely, when quantitative evaluations find bilingual education to be effective, critics of bilingual education are equally quick to point out that factors involved in the selection of children in the bilingual program and not the program itself may be responsible for the results (see Baker and deKanter, 1981).

Quantitative evaluations of bilingual education also require a degree of consensus concerning the goals of bilingual education. This may not pose a serious problem at state or Federal levels, since in the U.S. the primary objective of bilingual education is to facilitate the transition of limited-English-proficient children into all-English program as quickly as possible. However, many bilingual program directors, teachers, parents, and students may well have other goals, e.g., the development of a high level of proficiency in the children's native language and/or the maintenance of certain features and knowledge of their native culture. Therefore, quantitative evaluations designed with state and/or Federal reviewers in mind may not provide information relevant to the needs and concerns of people who are closer to the program. In this respect it is interesting to note that the current system being used in Illinois to collect evaluative information on state-funded bilingual programs does not involve the collection and reporting of any data on native-language skills (Illinois State Board of Education, 1981). In my experience I have found this rift between "official" goals and actual goals at the program level to exist primarily in the U.S., where bilingual education program directors and some bilingual teachers often have "maintenance" or "restorationist" goals with respect to children's native languages, while official state or Federal policy is oriented to the "transitional" goal of moving children from bilingual to all-English instruction as quickly as possible.

A final weakness of quantitative methods to evaluate bilingual education programs is the necessity of reliable and valid instruments to translate the behavior and feelings of those involved in the programs into meaningful, useful numbers. Among the many problems related to this requirement of a quantitative approach to evaluation are: (a) Deciding whether norm-referenced or criterion-referenced tests are appropriate (see Block, 1971; Ebel, 1971); (b) the possible bias of standardized tests for language minority students (see Olmedo, 1977); and (c) locating or constructing measures of language proficiency which

are practical to administer and yet take into account the various components of language skills which are now believed to make up communicative competence (see Canale and Swain, 1980). These remain serious issues in the U.S. in spite of the large amount of research undertaken to deal with them and the efforts of major publishers in the U.S. to produce reliable, valid and unbiased standardized tests of academic achievement. The reliance of quantitative evaluation approaches on reliable and valid measuring instruments means that this approach is of limited usefulness in settings where tests are either not available or where there are little or no resources available for the construction and validation of such tests.

Qualitative Evaluation Approaches. If quantitative approaches to the evaluation of bilingual education programs suffer from so many apparent weaknesses, might not more qualitative (or naturalistic) evaluation methods be better suited to the task? While much has recently been written about the need for and strengths of qualitative approaches to educational evaluation (see Guba's, 1978, naturalistic inquiry; Stake's, 1975, 1978, responsive evaluation and case study approach; Patton's 1980, qualitative-evaluation methods), there seems to have been very little use of these qualitative evaluation approachs to examine bilingual education programs in the U.S. or abroad. This is not surprising in the U.S. when one considers that Federal regulations require that evaluations of Title VII bilingual education programs employ an evaluation approach based on test scores and appropriate statistical analyses to show that the bilingual program is having an impact on participating students' academic achievement (see Burry, 1979, p. 11).

Qualitative evaluation approaches do offer some appealing advantages over more quantitative methods, although they have weaknesses of their own. One important practical strength of a qualitative approach to evaluating bilingual education programs is that random assignment of students to bilingual and non-bilingual programs is not

necessary, nor is it even necessary to include a control or comparison group in the evaluation design, since the primary purpose of a qualitative approach is to understand how the program works and how it is viewed by students, parents, and administrators involved in the program. Of course, a comparison or control group may be included in the evaluation design and may well provide important evaluative information, but the apparently unsolvable problem of random assignment or of assuring group equivalence before the program begins is not a prime concern.

Another strength of a qualitative approach to evaluation is that it requires that the evaluator obtain a detailed, first-hand look at the program in action, an experience not required by more quantitative evaluation approaches. Thus, while I spent considerable time observing literacy classes in the Southern Sudan and listening to and analyzing tape recordings of these classes, the classroom experience I obtained during the more quantitative evaluations of French immersion programs in Canada and the Yoruba project in Nigeria was limited to supervising the administration of group tests and courtesy classroom visits which involved greeting the teacher and students and perhaps observing the class in action for a very short period of time. While in-depth class observation may not be necessary if a program is known to be well implemented and is doing well by quantitative criteria (e.g., the French immersion programs in Canada), this type of information can be invaluable for recommending change if a program is not achieving its goals. Thus, while the reading test performance of children in project classes of the Sudan project clearly indicated that they were having considerable difficulty in learning to read, the classroom observation gives an indication as to why this was the case. As mentioned earlier, it was found that there were only rare instances of classroom activities which actually required some type of reading comprehension on the part of the students. This qualitative finding suggests obvious steps for the improvement of the project--steps that would not be suggested by quantitative evaluation methods alone.

The final advantage of qualitative evaluation methods to be mentioned here is that this approach does not require the availability of reliable and valid quantitative measures of the academic achievement, language proficiency, or attitudes of participating students. This is a particularly desirable feature in settings where such measures are not available (and, as mentioned above, it could be argued that they are not even available in the U.S.).

Unfortunately, qualitative approaches to evaluation suffer from a number of weaknesses. One of these is the apparently subjective basis (using Scriven's, 1972, qualitative meaning of this word) on which success or failure of a bilingual program is judged. This seems to be a particular problem in the U.S. where most evaluators of individual bilingual education programs appear to have pre-existing views of the worth and merit of bilingual education. It is imperative, therefore, that evaluators using qualitative approaches provide strong support for their conclusions by citing supportive evidence from as many different sources as possible. While this should be a particular concern of evaluators using qualitative evaluation approaches, it appears to be a widely unrecognized problem of quantitative approaches as well.

While the use of tests and statistical analyses appears on the surface to be more objective (and hence reliable), statistical tests in themselves do <u>not</u> tell us whether the difference found in performance on a certain measure between bilingually and monlingually instructed groups is in fact meaningfully significant. This is because inferential statistical tests are influenced by the size of the groups included in the evaluation, so that the same small difference between group means, which has no statistical significance when comparing small groups of students, may in fact be statistically significant when comparing larger groups. Unfortunately, few evaluators seem to take this fact into account and often consider all statistically significant differences to be practically significant (see Popham, 1975, p. 239). This obscures

the fact that all conclusions regarding the impact of a bilingual program, whether measured quantitatively or qualitatively, are essentially based on subjective judgments.

Another feature of qualitative approaches to evaluating bilingual pcgrams that can be considered a drawback within certain contexts is that the data which are collected (e.g., classroom observations, open-ended interviews with students, collections of students' written compositions) are usually difficult to reduce and summarize. While this same feature permits the diligent evaluation reader to get a good idea of a particular progam in operation, it makes it extremely difficult for policy-makers to review a large number of qualitative evaluations for the purpose of giving an empirical base to policy decisions. For this reason it might be argued that while qualitative evaluations are more appropriate for obtaining and disseminating detailed information about a particular program and for providing information for program changes, quantitative evaluation approaches are more appropriate for the synthesis of a large amount of data obtained from a large base for the purpose of making policy decisions. Consistent with this observation is the fact that both the AIR evaluation (Danoff, 1978) and the recent review of bilingual evaluations prepared for the Office of Planning and Budget of the Department of Education (Baker and deKanter, 1981) were concerned exclusively with quantitative evaluation methods and criteria. It would appear, then, that evaluators using qualitative approaches to evaluating bilingual education programs run the risk of having their work ignored by reviewers and policy planners.

Finally, it must be mentioned that the use of qualitative evaluation techniques typically demand a very large amount of time and effort for data collection, analysis, and report writing. Also, qualitative methods are difficult, if not impossible, to employ in situations where the evaluator does not have a thorough knowledge of the language and culture of the school setting unless trainable research assistants can be found.

Combining Quantitative and Qualitative Approaches to Evaluation

After having considered some of the strengths and weaknesses of both quantitative and qualitative approaches to the evaluation of bilingual education programs, it would appear that an effective evaluation approach would be to combine aspects of both approaches in order to minimize the weaknesses and maximize the strengths of the overall evaluation methodology. Although this is certainly not a new proposal, it is a fairly recent one (see Patton, 1980; Fry et al., 1981) and it constitutes an approach which apparently has not been much used for the evaluation of bilingual education programs. Although my work in the Southern Sudan attempted to use such an approach, it must be admitted that the relatively short amount of time devoted to class observations (ranging from about one to three hours for each classroom included in the evaluation) did not permit the type of rich, ethnographic data collection and interpretation which is characteristic of qualitative evaluation methods. Unfortunately, while such a combined approach might look appealing on paper, there are probably very few educational evaluators working today who have the necessary expertise to carry out such an evaluation, particularly of bilingual education programs. It also seems that an evaluation of a bilingual education program which attempts to combine aspects of both quantitative and qualitative evaluation approaches would have to be very carefully designed if it were to capitalize on the strengths of both approaches and not suffer from their combined weaknesses.

While it is beyond the scope of this paper to go into details of how these two approaches can be combined for evaluating bilingual education pograms (see Fry et al., 1981 and Patton, 1980 for suggestions on how this can be done for general evaluation and social science research), it is hoped that this paper has provided a strong rationale for such an approach. If our evaluations of bilingual education are to be useful in detecting impact of such programs, disseminating

information on how such programs operate, and providing an empirical base for improving bilingual education, then it is obvious that currently used evaluation approaches are not meeting these needs. New, innovative evaluation approaches are needed. Fortunately, many innovations in educational evaluation have occurred over recent years which have much to add to our work in bilingual education. It is our challenge to make the most of them and to continue to develop new evaluation techniques appropriate for determining the most effective programs for educating language minority student in the U.S.

References

Afolayan, A. The six-year primary project in Nigeria. In A. Bamgbose (Ed.). Mother tongue education: The West African experience. London: Hodder and Stoughton, 1976.

Baker, K. A., & deKanter, A. A. Effectiveness of bilingual education: A review of the literature. Washington, DC: U.S. Department of Education, Office of Planning, Budget and Evaluation, 1981.

Block, J. H. Criterion-referenced measurements: Potential. School Review, 1971, 79, 189-298.

Burry, J. Evaluation in bilingual education. Evaluation Comment, 1979, 6(1), 1-14.

Cowan, J. R. Report on a reading workshop held for Summer Institute of Linguistics/Institute of Regional Languages Personnel in Maridi (Sudan): October 1980. Unpublished manuscript, December 1980. (Available from Center for Applied Linguistics, Washington, DC).

Cziko, G. A. Vernacular education in the Southern Sudan: A test case for literacy. Journal of Multilingual and Multicultural Development, in press.

Cziko, G. A., & Ojerinde, A. Yoruba six-year primary project: The June 1976 evaluation. Unpublished research report, November 1976. (Available from Institute of Education, University of Ife, Ile-Ife, Nigeria.)

Danoff, M. N. Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program: Overview of study and findings. Palo Alto, CA: American Institutes for Research, 1978. (ERIC Document Reproduction Service No. 154 634)

Ebel, R. L. Criterion-referenced measurements: Limitations. School Review, 1971, 79, 282-288.

Epstein, N. Language, ethnicity, and the schools: Policy alternatives for bilingual-bicultural education. Washington, DC: Institute for Educational Leadership, 1977.

Fry, G., Chantavanich, S., & Chantavanich, A. Merging quantitative and qualitative research techniques: Toward a new research paradigm. Anthropology and Education Quarterly, 1981, 12(2), 145-158.

Genesee, F. The suitability of immersion programs for all children. The Canadian Modern Language Review, 1976, 32, 494-515.

Gray, T. C. From the Center for Applied Linguistics. The Bilingual Journal, 1978, 2, 23-24.

Gray, T. C. Current Status of Legislation on Bilingual Education. Arlington, VA: Center for Applied Linguistics, 1981.

Guba, E. G. Toward a methodology of naturalistic inquiry in educational evaluation (CSE Monograph Series in Education No. 8). Los Angeles, CA: Center for the Study of Evaluation, 1978. (ERIC Document Reproduction Service No, ED 164 599)

Illinois State Board of Education. The first annual program summary and evaluation report of transitional bilingual education programs in Illinois: 1979-1980. Springfield, IL: Author, 1981.

Lambert, W. E., & Tucker, G. R. Bilingual education of children: The St. Lambert experiment. Rowley, MA: Newbury House, 1972.

Ojerinde, A. The use of a mother tongue, Yoruba, as a medium of instruction in Nigerian schools (Doctoral Dissertation, Cornell University, 1978). Dissertation Abstracts International, 1979, 39, 4202A.

Olmedo, E. L. Psychological testing and the Chicano, In J. L. Martinez, Jr. (Ed.). Chicano Psychology. New York, NJ: Academic Press, 1977.

Patton, M. Q. Qualitative evaluation methods. Beverly Hills, CA: Sage, 1980.

Popham, W. J. Educational evaluation. Englewood Cliffs, NJ: Prentice Hall, 1975.

Scriven, M. Objectivity and subjectivity in education research. In H. B. Dunkel, et al. (Eds.). Philosophical redirection of educational research. Chicago, IL: National Society for the Study of Education, 1972.

Stake, R. E. Program evaluation, particularly responsive evaluation (Occasional Paper No. 5.). Kalamazoo, MI: Western Michigan University, November 1975.

State, R. E. The case study method in social inquiry. Educational Researcher, 1978, 7, 5-8.

Troike, R. C. Research evidence for the effectiveness of bilingual education. NABE Journal, 1978, 3, 13-24.